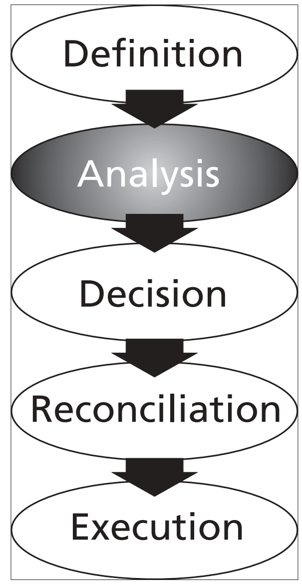
CHAPTER 7

ANALYSIS CONCEPTS: MODELING



Farming looks mighty easy when your plow is a pencil,
and you're a thousand miles from the cornfield.
-Dwight D. Eisenhower, speech in Peoria, Illinois, 25 September 1956

MODELS, SIMPLE OR COMPLEX, are the abstract constructs we use to compare alternatives. In defense resource allocation, models have four functions: organization of a problem, comparison of alternatives, measurement, and prediction. The first function we discussed in Chapter 2 as we defined and organized the problem. This chapter expands on the methods for combining criteria into models specifically designed to support analysis. The most frequent use of analytic models in DoD is to compare procurement and policy options on the basis of cost and effectiveness in the Analysis Phase.[1]

Our intention in this chapter is to familiarize you with the analytic modeling tools and terminology of the analyst, not to have you memorize classifications and characteristics of models. As you read this chapter, remember our goal is to make you a critical director and consumer of analysis who can confidently evaluate modeling proposals. By understanding the difference between good models and bad models and by subjecting analytic models to professional scrutiny in terms of validity, reliability, and practicality, you will be able to evaluate the quality of analysis without becoming a subject matter expert and thereby make good executive decisions.

Characteristics of Analytic Models

Analytic models are a specific class of models. They are so named because they are models composed of the separate parts of a problem—a problem identified by the analytic objective and the parts that were important enough to be facts, assumptions, or criteria. Analytic models require that their builders and users have an understanding of how those parts fit together. Analytic models are, at heart, based on the scientific method and they have a clear logical or mathematical structure.[2]

1. Within DoD, some cost and effectiveness analyses are given names that specify their structure and content. DoD uses the Analysis of Alternatives format, which superseded the Cost and Operational Effectiveness Analysis format in 1996, to support acquisition milestone reviews. DoD initiated the V-22 case study used throughout this text as a Cost and Operational Effectiveness Analysis.
2. Mathematics is often the language of modelers because of its wide applicability to seemingly unrelated problems. For example, the same form of equation describes the decay of a radioactive isotope, the swing of a pendulum, the decline of a population, etc.

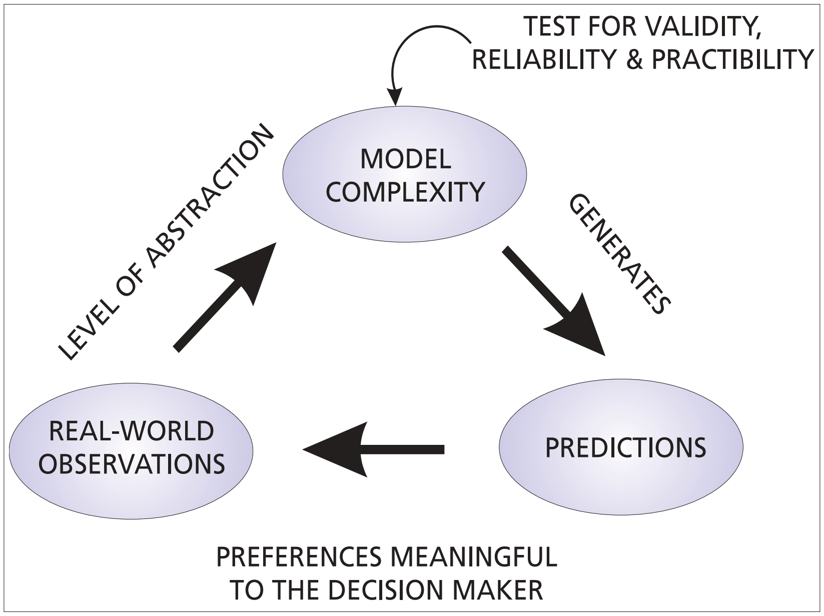


Figure 7-1. Models and the Scientific Method.[3]

With unlimited resources, including time and money, we would not need models, and we could satisfy our requirements with full-sized experiments and real-world observations. But we are forced to model by the prohibitive cost of experimenting in the real world, although the real world is our starting point for the analytic model as we show in figure 7-1. We will examine the major characteristics of analytic models—abstraction, complexity, and prediction—in more detail below. When we defined the problem, we set the stage for selecting many of the characteristics of our analytic models.

LEVEL OF ABSTRACTION

Our translation of the real world into an analytic model is an abstraction because it reflects a simplified reality containing only those factors and relationships we deemed important for solving the problem. The level of abstraction is inversely proportional to the degree to which the model literally replicates reality. Some models have a great physical resemblance to reality, e.g., mock-ups, prototypes, and miniatures. Others models, like the differential equations that represent the airflow around a ballistic missile warhead, bear little resemblance to the physical world they are used to investigate.

Most of the analytic models we use in DoD vary greatly from reality because they are based on mathematics or use scale representations such as time compression. Full-scale models, such as prototype aircraft, have few (if any) departures from reality. Policy analysis also takes advantage of full-scale models; before launching a new quality of life program, we usually test the policy in a pilot program as we see now being done with several military health program initiatives.

Iconic models are scaled down replicas of the real world, such as model airplanes, maps, globes, and photographs. We use iconic models to provide information without going to the expense or difficulty of building full-scale models, assuming the model's performance mimics real world performance. Note that the level of abstraction need not be connected to the complexity of the model. Some highly abstract models are very complex, such as those for space flight planning, and other highly abstract models are quite simple, like a flow chart.

COMPLEXITY

The complexity of analytic models is a function of the number of variables we need to measure, the resolution to which we measure them, and how many resources we have available to model—validity, reliability, and practicality concerns. We will also have some uncertainty about how well the interactions among elements of the model reflect reality. In the simplest case, we have rigid, full-scale analytic models; in the most complex models, we include interactions among the variables and insert events while the model is running.

3. Adapted from Samuel B. Richmond, *Operations Research for Management Decisions,* New York: The Ronald Press Co., 1968:.30.

Static models are representations of reality at a fixed point in time, freezing both time and position, such as a map or an organization chart. Because they do not incorporate change, they are generally simple and inexpensive. Their simplicity is helpful when we have a wide divergence of opinions and perceptions about the problem amongst the decision participants.

Dynamic models incorporate change in terms of time, events, and motion, e.g., a fuel usage curve that displays gallons consumed as a function of hours in operation, a graph of accidents as a function of crew rest, and computer simulations of weather conditions. Adding change to create dynamic models adds complexity and uncertainty to them. Dynamic models are often more difficult to describe and display, especially the interactions among variables. Even so, we use them to take into account interactions that we know to be important in the real world. To the extent that they succeed, dynamic models reflect the real world better than static models.

Yet, even simple static models may include important uncertainties; cartographers do not survey every square inch of the terrain represented by a map, yet every square inch is represented. Just as in the Definition Phase, the analysts and we are forced to make modeling assumptions to cope with uncertainty. We must take into account the uncertainty of important but uncontrollable variables in dynamic and interactive models, such as weather or the price of fuel.

NETWORKS, COMPLEXITY, AND UNCERTAINTY

Predictive modeling is based on the assumption that the future can, at least in part, be forecast by knowing the past and understanding how variables, including our criteria, act and interact within their environment. Some prominent theorists believe this assumption is fundamentally wrong. Chaos Theory is a well-known approach that describes a world where chance rules supreme and confounds our ability to predict outcomes that may vary wildly despite nearly identical initial conditions.[4] When we look at initial conditions and then outcomes long afterward, it is very difficult to identify exactly why the outcomes were so different. If, however, we start at the beginning and catalog the intervening events with ever more resolution, we can identify a linear series of decision points and chance occurrences (nodes) that keep branching out until we have a huge but exhaustive set of possible outcomes.

As we progress from node to node, some branches may merge into nodes with other branches, creating multiple paths to the same outcome—a network. The path we uncover by reverse engineering the outcome is one possible path among many in a network of unknown dimensions. Chaos theorists see any progression of events to an outcome as non-unique; one path along a network may be repeated later, but neither the path nor the outcome is predestined by the initial conditions. The longer the time interval and the more numerous the events, the larger and more complex the network and collection of paths and outcomes become and the more difficult it is to model. We can complicate the network further by adding more starting points.

Chaos Theory operates from the assumptions that: (1) the future is not linked to the past in a linear fashion, therefore we need higher order mathematics to approximate or model future behavior; (2) events in nature are very sensitive to initial conditions, therefore small, hardly measurable changes in one variable at the beginning of a chain of events can dramatically change the

4. A typical example posits a child dropping two ping pong balls into the Niagara River above the falls. One winds up washing ashore near the base of the falls and the other comes to rest on the coast of Africa.

overall outcome; and (3) stability in a chaotic system is unnatural and quite temporary, but where stability does exist it is determined by the relationship of very few variables. Therefore, controlling these key variables can control behavior in a chaotic system for a brief period. But, because this controllable time period is brief, we cannot predict the distant future with any degree of certainty.

The global weather system is an excellent example of a chaotic network. We cannot predict the weather accurately more than 72 hours in advance; our attempts to predict the weather further beyond the simplest generalities are futile, according to chaos theorists. But we know the probable range of outcomes from the global weather network, therefore civil engineers can plan their designs around 50-year storms, i.e., severe storms that statistically happen every 50 years, while no one tries to say exactly when the next one will occur. Also, we may be able to control the weather over a short period if we could identify and learn to manipulate the key parameters, such as by seeding clouds to precipitate rain.

These ideas can have important implications for the study of war. The network model is much more compatible with our experience of war than the chessboard. Analysts are not able to predict other than the grossest outcomes of war. If we can identify and learn to man-age the key determinants of the outcome of the process (which may be very few), then we can control the process of war over short, critical periods. This requires that we use higher mathematics and probability, accepting ranges of outcomes like worst case, best case, and most likely case to compensate for the much higher levels of uncertainty we will have to accept with network modeling.

PREDICTION

Analytic models make predictions about the outcomes we should expect; given our decision to use a particular model, our choice of input values, and our choices between alternative courses of action. If a decision-maker has confidence in a model and in the chosen set of input values, these predictions will help him choose a course of action.

Whenever we can, we evaluate a model's quality by comparing its predictions with real-world outcomes, then we calibrate it to better predict and improve our confidence in it. Of course, the extent to which we can do this depends on the kind of problem we are investigating. Certain problems make it relatively easy to test model results against real-world outcomes (e.g., how fuel consumption varies as a function of the kind of flight training we are doing).

The more the problem we are investigating involves predicting results in combat, the harder it will become to test model results against real-world outcomes. For one thing, we have a small number of real-world wars against which to compare our model results. In addition, careful historical analysis of actual battles shows that outcomes depend on a series of hard-to-replicate and unlikely-to-recur particular events.

Even if it's hard to know if a particular model is doing a good job of predicting combat outcomes, we can learn a great deal from modeling. For one thing, building a model forces us to say what premises we have to believe, in order to believe a particular prediction. Sometimes we can subject those premises to empirical tests. Depending on what those tests show, we can revise our prediction and, ideally, get closer to understanding "ground truth."

PREDICTION AND THE MOBILITY REQUIREMENTS STUDY AIRLIFT MODELS

In 1991, the Joint Staff evaluated U.S. strategic lift to determine whether it was adequate to deploy U.S. forces in time to achieve national military objectives. The abstract models the Joint Staff used for comparing alternative aircraft fleets for strategic airlift had simple criteria and algorithms. Their criteria included gross weight carried, airspeed, mechanical reliability, and range; it did not include outsize or oversize cargo.[5] Either of these types of cargo causes the U.S. Transportation Command to use aircraft contrary to the model, i.e., the transport aircraft cannot load cargo to the study capacity. Thus, the models did not accurately predict the actual behavior of the strategic airlift fleets under true operational conditions.

Despite their limitations, the models were still very useful for the Joint Staff. They enabled them to determine the relative differences in performance among the aircraft fleet alternatives. The Joint Staff did not mistake these insights into relative performance differences among fleets for absolute outcomes. While they could conclude one airlift fleet had 30% more capacity than another did, they knew they could not say that the first fleet would deliver X tons of supplies in seven days while another took ten. Although we would like the model to predict faithfully what will happen, we can often settle for models that show differences in relative performance, despite their inability to evaluate absolute performance.

Models differ in their ability to predict what will happen in the real world. Some models do not predict the absolute outcome of events very well, but they are still useful as long as they display a relative difference in performance among the alternatives that will carry into the real world.

Types of Analytic Models

DoD uses many standard models for analysis. For example, Joint Simulation System (JSIMS) provides a validated computer-simulated environment for use by the CINCs, their components, other joint organizations, and the Services to jointly educate, train, develop doctrine and tactics, formulate and assess operational plans, assess war-fighting situations, define operational requirements, and provide operational input to the acquisition process. Another example is actually a suite for four simulation models, JQUAD, which contains electronic warfare, command and control, network, and operational intelligence models. These models, along with numerous others that have been validated by the Pentagon, establish methods for the most frequent analyses by using common frames of reference. Using an already-accepted model automatically focuses discussion on the unique aspects of the decision whereas with a new model, we will have to gain acceptance before we can advocate our preferred alternative. Therefore, we should always consider modifying existing models to fit our decision rather than building a new model from scratch.

Below we list some of the more common types of analytic models that can be used for defense resource allocation decisions. Which model we select depends entirely upon the situation; an appropriate fit between model and problem is paramount. Because models vary in abstraction, com-

5. Outsize cargo, e.g., tanks, exceeds 9.75 feet in width, 8.75 feet in height, or 90.8 feet in length; it is the largest class of air cargo and it fits into C-5 and C-17 aircraft but not C-141s. Oversize cargo is typically a single item, like a pickup truck, that does not exceed the size of a standard 463L pallet but does not allow the air-craft to stow cargo to its maximum capacity or efficiency. C-141s can carry oversize cargo. Source: Military Airlift: Airlift Planning Factors, AFP 76-2 (C-1), 1982, p. 4-5.

plexity, and their ability to predict in different situations, we must have a clear problem definition and a thorough understanding of how our criteria interact before we select a model.

We often run models using scenarios as backdrops. Scenarios are situations, a collection of boundaries, including facts and assumptions from the Definition Phase, and other necessary conditions for running the model, such as location, time frame, sample size, etc. We may specify scenarios for the problem we are solving or have the analysts develop them based upon existing or predicted scenarios, e.g., the Defense Planning Guidance includes two appendices of illustrative scenarios (one current and one future) for force structure planning; the Combatant Commanders test their concepts of operations in scenarios loaded into large models.

Again, the names of these models are less important than understanding their character and understanding how we can apply them to different types of decisions. We also present them here because analysts often use this terminology in their descriptions and proposals.

DETERMINISTIC MODELS

These models require a thorough understanding of causes and effects in the environment or problem we are modeling. We change one or two key input variables, leave the other variables stable, and produce an outcome resulting from the input changes: input ***a*** leads to output ***b***. We use deterministic models when accurate prediction is especially important and we have a high level of certainty about the controlled variables in the model.

Many simulators use deterministic models. In an aircraft flight simulator, moving a control in a particular manner causes change in the flight characteristics related to it. The model determines the overall effect the control adjustment will have, and reacts accordingly. Deterministic models, assuming they are built correctly, are very reliable predictors—they will produce the same result under the same circumstances every time. Therefore, we must decide if that is also true of the portion of the real world we are trying to describe before we commit to a deterministic model.

INVENTORY MODELS

Used primarily by logisticians to manage stock levels, these models play an important role in force planning, particularly in procurement, because life cycle costs are dramatically affected by spare parts and energy consumption: their cost, usage rate, storage, and delivery. To be effective, these models require solid estimates about user consumption. Generally, inventory models contain two or more competing cost curves, e.g., storage cost and transaction cost for spare parts.

Using a naval example, storage cost is the expense of maintaining an inventory of spare parts for rapid issue to the Fleet. Transaction cost is the cost of obtaining an item directly from a supplier on demand; generally this takes longer than an internal transaction within the Navy and is more expensive because there are no price breaks for large volume purchases. But if we store too many of these spare parts, we have several problems. First, the Navy may have too much purchasing power tied up in inventory—stocking the inventory imposes opportunity costs in other areas. Second, warehousing them creates costs by itself. Finally, if these parts are technologically perishable, we will waste resources if they are never consumed and they have little disposal value. The analyst seeks to find the lowest cost over the life cycle of the system to balance the two costs and recommend an inventory level to the Navy that optimizes responsiveness (adequate inventory within the Navy) and transaction costs (frequency of replenishment of that inventory).

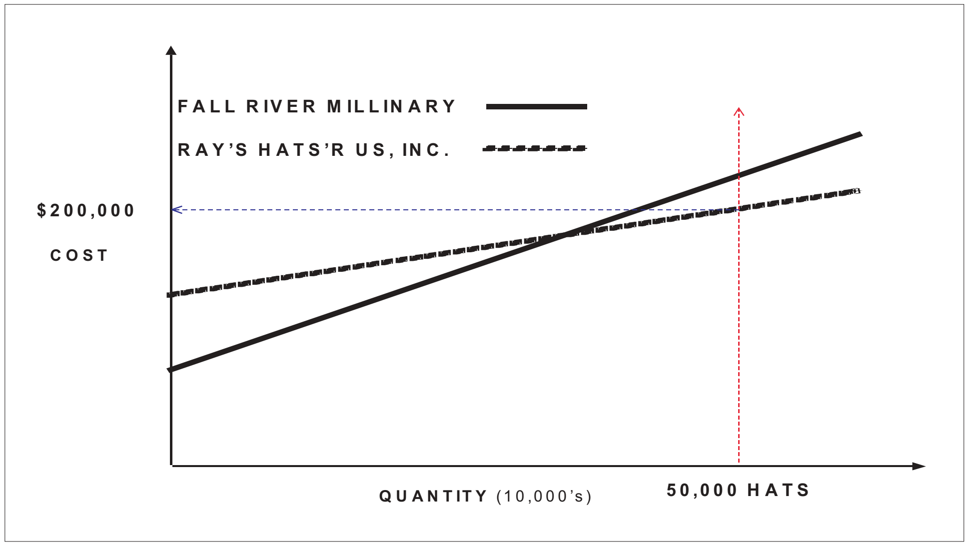


Figure 7-2. Allocation Model.

ALLOCATION MODELS

Allocation models examine the most efficient assignment of resources to tasks. Typically, we use spreadsheet programs to explore the effects of a change in one area upon another. In DoD, we use allocation models to solve assignment problems wherein we have a number of tasks and a number of units that can fulfill them. When a CINC provides guidance for a quarterly schedule for his ships, he considers current operational requirements, exercise participation, in port maintenance requirements, and port visits in support of his Theater Engagement Plan. These requirements and the list of ships available could be built into an allocation model that would optimize scheduling, or at least provide a rough schedule to use as a starting point. Allocation models are also useful for solving transportation or network problems in which the analyst seeks the most efficient path from a starting point to an outcome. The variables in the model behave very much like the values in utility curves (see Chapter 6).

An example of an allocation model is shown in figure 7-2. In this case, we are buying headgear to stock a uniform store and we have two competitors that have provided price curves. We need 50,000 hats (fixed effectiveness), so we select the most optimal solution by reading up and then across to identify the lowest cost—Ray's at $200,000.

STOCHASTIC MODELS

Stochastic models are always dynamic or interactive; they incorporate time, randomness, and probability theory. They are very useful when we have high degrees of uncertainty, when input ***a*** yields output ***b, c, d,*** or ***e***—or any combination of them. One branch of stochastic models involves queuing processes. Queuing models derive their name from their initial applications in the service sector, i.e., they were used to identify the number of passenger gates and their arrangement in airports. To build this kind of queuing model, the analyst first in-puts the problem boundaries: the service or process time (fixed and known) and the behavior rules for processing the people in line, including decision rules such as: First In, First Out; Last In, First Out; or Very Important People To The Head Of The Line. Then the analyst designates the number of service stations (the range of solutions) for different runs of the model. The model uses stochastic methods to input customer arrival times (the random or probabilistic event) with a variety of surges and slack periods (random or designed by the flight schedule) as the model runs. The output of each run is information about customer waiting times: average, longest, mean, etc. An airline using this model could set a goal for an average waiting time and then use the model to predict how many customer stations it needs manned to satisfy loading at different times for different days.

In DoD, queuing models help us plan the overall capacity we require for maintenance and support of a force structure. To support the 1997 Quadrennial Defense Review, DoD constructed a wargame called Dynamic Commitment to examine the demands that might be placed against the U.S. military in the next fifteen years. The analysts constructed a series of scenarios from major theater wars to a variety of peace operations and let the model generate their sequencing stochastically in accordance with some rules, i.e., no more than two major theaters wars simultaneously and at least five years between major theater wars in the same theater. None of the strings was meant to be a literal prediction of the future; rather the analysts used the results of their many runs to identify the character of the force structure that was most likely to be successful at meeting every requirement. (Unfortunately, the force structures identified exceeded the resources likely to be available by a wide margin.)

Markov chains are another stochastic modeling tool. Markov chains exist when the probability of one event happening depends on what happened in the event that immediately preceded it. These are the mathematical equivalents of the branches and sequels we use in operational planning. For example, service-recruiting targets are a function of force structure requirements and retention; changing either will affect recruiting goals. Stochastic modeling has become prevalent with the use of computers that can manipulate a plethora of data, equations, alternatives, events, and possible outcomes; therefore we use these models to support wargaming.

COST ANALYTIC MODELS

Cost models range from the very simple to the extraordinarily detailed. Some use advanced mathematical techniques, others only basic arithmetic. Some require extensive computer support, others analysts build manually or with simple spreadsheets. Remember that cost estimating methods tend to overlook costs that cannot be measured in dollars and these other types of cost are often more important to us than dollars alone.

For existing weapons and support systems, we can estimate cost using historical data. However, for many force-planning decisions, the systems do not yet exist. Fortunately, there are numerous cost estimating methods that can be used to predict future costs. Three of the most common are the analogy, parametric, and industrial engineering methods.

Analogy Method

When detailed cost data is not available, an analyst may estimate cost by making direct comparisons with similar existing systems. For example, using the analogy method, we can approximate the value of surplus land on a DoD installation based on the sales of similar property nearby. We often estimate low-value equipment proposals, commodity purchases, and operating and support expenses using analogies. This method is also very effective for estimating the cost of off-the-shelf equipment where comparable prices are as close as the nearest catalog. In order to use the analogy method for new or complex concepts, an analyst needs considerable expertise and judgment. The less compatible the subject and the model, and the older the existing comparator, the less confidence we have in this kind of cost estimate.

Parametric Method

We may deem it impossible to find an appropriate analogy to use to estimate cost for a new item. However, we may be able to identify characteristics or parameters of the new system that are similar to the characteristics of other existing systems. Using those carefully identified parameters, we seek a cost estimating relationship that we can project onto the new acquisition. The cost estimating relationship sets this method apart from the analogy method. It is a mathematical expression that relates one or more particular acquisition characteristics to cost, e.g., cost per ton for the construction of a ship. Note that the cost-estimating ratio itself may be based on an analogy; we may estimate the cost of a new government warehouse, larger than any previous building we have contracted, by multiplying the area times the cost per square foot of an airplane hangar or large civilian warehouse.

We use the parametric method in DoD for estimates early in the Defense Acquisition System. Parametric estimates can be very accurate when they are based on realistic, historical experience, as demonstrated in the accuracy of F/A-18E/F cost estimates, which were based on the costs of the C/D model. Moreover, we can calculate the cost estimate quickly once we establish the cost estimating relationship. Parametric costing may result in pessimistic estimates if we do not adjust the formulas based on historical experience for improved production methods or recent lessons learned.

Industrial Engineering Method

The industrial engineering cost estimating method is often referred to as the bottom up approach. An analyst using this method consolidates estimates for various segments of a project into a total estimate for the entire project. Government analysts estimating the cost of a new building use this method by estimating the structural, electrical, plumbing, heating and air conditioning, and other component costs of the projects. They may break each of these estimates down further into sub-components such as labor, materials, equipment, etc. The industrial engineering method is the most thorough way of estimating cost, but it can be quite time consuming.

Evaluating the Model

Before the analyst runs the model and we compare alternatives, we will evaluate the model to ensure it reflects how we think the criteria behave and interact. First, we review the Definition Phase to ensure the guidance we gave the analyst conforms to our analytic objective and that our analytic objective still makes sense. Then we review the analyst's model proposal to ensure it aligns well to the analytic objective, e.g., we do not want to use a complex stochastic model to evaluate a simple decision about bulk commodity purchases. This kind of mismatch happens most often when we use an existing model for a new decision situation. Then we evaluate the model's level of abstraction, complexity, and predictive qualities in terms of validity, reliability, and practicality. When we are satisfied with the qualities of the model, we should obtain the decision maker's approval before proceeding further.

MODEL VALIDITY

As we examine the validity of our model, we ask whether it captures the most important behaviors of the alternatives at the right level of resolution—does it model the right things? Do the criteria reflect our perceptions of reality? In a weighted model, do our utility curves and weights reflect our values? The boundaries in the model must be consistent with the elements we identified in the Definition Phase. It must model the alternatives objectively. We must understand the predictive qualities of our model to ensure it helps us distinguish among the outcomes and we must have confidence that the models' projections are consistent with the real world. Finally, the model's level of complexity must be appropriate for the decision maker.

We need to view the model as a totality, also. We can get mesmerized by the detailed evaluation of criteria to a point where we lose sight of the analytic objective. Air campaign planners, used to trading off strengths and weaknesses of tactical aircraft, sometimes need to be reminded to use models that are robust enough to include tactical missiles, bombers, and attack helicopters.

MODEL RELIABILITY

Where reliability is concerned, we are interested in the model's behavior: does it measure well? The internal consistency of our model determines whether we are confident that the results of the model (predictions) will be the same whenever the model is used under similar circumstances. We must be able to measure the criteria well at the model's level of abstraction.

The model may be affected by measurement errors as we collect data on the criteria, especially if we fail to measure with enough resolution. We must determine how much error is tolerable while running the model and whether we must measure some criteria more precisely than others. If others are providing data, we must be satisfied with its accuracy. Generally, the more abstract the model, the simpler it becomes and the more forgiving it is of our measurement errors.

MODEL PRACTICALITY

Some models are more costly than others, and we seek to balance realism (validity and reliability) with cost as we address the model's practicality. Reducing cost to avoid the difficulty and expense of real world testing is our reason for modeling in the first place. The resources we consume in modeling should be commensurate with the importance and urgency of the problem to our organization.

The more abstraction we accept (the further we move away from reality) in the model, the more vulnerable we are to criticisms that the model does not reflect the real world. In addition, our results are more difficult to "prove." If the model's predictions are too unreliable, we will have to improve its data, reduce its level of abstraction or make the model more complex. We add complexity most often by making the algorithm more intricate, by adding variables (not necessarily criteria), or by increasing the level of detail in their measurement. All of this takes resources—time and money.

Analytic Models and the Information Age

The ability to store and retrieve data electronically from sources all over the world has greatly improved the quality of analysis in general and models in particular. Their validity and reliability are increasing as computers allow increasing complexity without degrading reliability significantly and at a reasonable cost. But computers may also conceal errors if we fail to understand the assumptions made by programmers and how they related our criteria to one another. “Garbage in, garbage out,” requires we be able to identify what is garbage.

DECISION SUPPORT SYSTEMS

Decision support systems are interactive software we run on computer hardware ranging from mainframes to networks to laptop personal computers. Decision support systems are very useful for organizing and manipulating subjective inputs from multiple participants in a decision and converting them into preferences for alternatives. The simpler systems help us build weighted models to compare procurement alternatives; the more complex decision support systems help us make force structure and policy decisions.

Decision support systems allow us to introduce structure and rigor to very complex problems and they are especially valuable when we cannot adapt other techniques to model the problem. For example, the Decision Strategies Department of the Naval War College, which has professional facilitators and uses a network of laptop computers, has examined policy alternatives for issues like confidence-building measures between Greece and Turkey and NATO enlargement. Because of the constant requirement for subjective judgments, we strongly desire the decision maker to be present when we use a decision support system for a policy issue.

NETWORKED MODELS

Our ability to exchange data through computer networks makes data commonality feasible and the process of data collection much easier. Decision makers in different locations can view the same spreadsheets and do sensitivity analysis during a teleconference from their workstations. The Joint Strike Fighter program is using this shared data base capability. The contractors and the program office use a common cost model; the DoD Program Manager can discuss cost data with a contractor while they both view the same database, tremendously simplifying coordination and reporting.

Using the Model to Evaluate Alternatives

Once we are satisfied with the model, we insert the alternatives and evaluate each. Recall that sometimes we have the alternatives before we build the model. In this case, we may have tailored our model to highlight the differences between the known range of alternatives and our foreknowledge may affect our criteria selection in particular. Because we use criteria to discriminate among options, we are unlikely to select an attribute whose value is equivalent for each alternative as a criterion. As we run the model, however, new alternatives may emerge and that may require us to re-evaluate our criteria and adjust the model.

When we create or learn of the alternatives after we build the model, the application is more straightforward. Sometimes, however, an unusual alternative arises after we have assembled our model that forces us to reexamine it, either to add new criteria or to identify a new requirement we need to apply to all the alternatives. The new criteria may not have discriminated among the previous options because they scored similarly. The new requirement may be necessary to exclude impractical solutions, e.g., a training range may be ideal in every regard except it is too far from homeports.

After the model runs, we have its results. Depending upon the nature of the problem and the model we used, they may vary from identifying a single preferred option to a hierarchy of scores for different alternatives, or a series of tables. In any event, we should be able to interpret them easily and explain them to others with clarity as we did with the radio example in the previous chapter and which we will continue below. We should not hesitate to stop and examine the model if its results defy easy explanation. While the possibility exists of new and exciting insights, it is more likely we have made a mistake and we need to find it and correct it.

When we are satisfied with the results, we need to create reports and briefings to support the decision maker. The seniority of the decision maker, the time available for briefing, and the magnitude, urgency and importance of the problem we identified in the Definition Phase will determine the amount of detail we present. Naturally, we should be able to explain the connective tissue from the most general of slides down to the measurement data if need be, just as an Executive Summary derives from a formal report and the report is based on modeling and data (often included in appendices).

Sensitivity Analysis

After the analysts run the model and results emerge, we often observe that some facts, assumptions, or criteria have an unusually strong influence on the outcome. Also, the analysts may not have data for some variables when they run the model, so they assign them arbitrary values, effectively making their own assumptions. We need to know how sensitive the results of the analysis are to changes in the values of variables, particularly if those values were estimated. To establish how changes in the value of a particular variable affect outcomes, we fix the values of all the variables in the model except the one under study. We then run the model several times, using a different value for the variable under study—high, low, and medium values for example—to see how changes in that variable affect the results. This process is called sensitivity analysis.

We may use sensitivity analysis in many ways during the Analysis Phase. First, we may change the boundaries of the problem or the initial conditions by altering facts or assumptions. For example, during Dynamic Commitment, changing the scenario queue to allow only one major theater war at a time results in a significantly smaller force structure set. We may also directly change the weights in a weighted model or the values of a criterion for different alternatives to explore variants and combinations of options. We can use sensitivity analysis to examine a criterion through the estimated range of its measurement error to see if we need better data.

Computers enable us to conduct a vast amount of sensitivity analysis rapidly and easily. We can vary almost any data or assumption in the model to determine whether changes are important to the results. In addition to its information value, sensitivity analysis is a powerful cost saving technique. For example, one of the variables in the model may be very difficult and expensive to measure. If we establish a range of probable values, run the model, and the preference between the alternatives does not change for these different values, the model is insensitive to that variable and we can use an assumed value without undermining the analysis.

If the outcome does change with different values, it is sensitive to that variable and we need to find a way to measure it directly or through a proxy. If the sensitive variable is an assumption, our last resort may be to display multiple sets of results for the different values of the assumption. For example, if we are comparing the life cycle costs of aircraft alternatives, each with different fuel consumption rates, the relative difference among options may be sensitive to our assumed price of jet fuel. We can check for sensitivity by running the model with our lowest estimated fuel cost and again with our highest estimated fuel cost to see whether the cost rankings of the alternatives change.

Another way to employ sensitivity analysis is to change the weights in a weighted model (without changing the values or scores of any alternative's criteria) to see if a change in weight alone changes the relative rank order preference of the alternatives. For example, let us return to last chapter's portable radio scenario and the output of its weighted model (see figure 7-3 on next page). We reproduced the alternatives and criteria, with their weights added in parentheses, in table 7-1.

RADIO	PURCHASE COST (25)	TOC (15)	RANGE (15)	WEIGHT (20)	SECURITY (15)	RELIABILITY (10)	TOTAL (100)
POPIEL 1995	23	15	0	20	0	0	58
WHAMMO 3000	11	7	12	16	15	6	67
ZONKER 101	4	14	15	6	15	10	64

Table 7-1. Radio Alternatives.

Recall that when we ran our weighted model, the Whammo 3000 scored the highest. The Popiel 1995, a lightweight inexpensive option with the minimum acceptable performance, scored poorest. What happens to the results of using the model if we change the weights to reward cost and effectiveness equally, i.e., weight each 50 vice the original 40 and 60, respectively? Assuming the weight changes are spread proportionally down to the criteria in the lowest tier and the performance of the alternatives does not change, we get a new preference for the Popiel 1995 as shown in table 7-2:

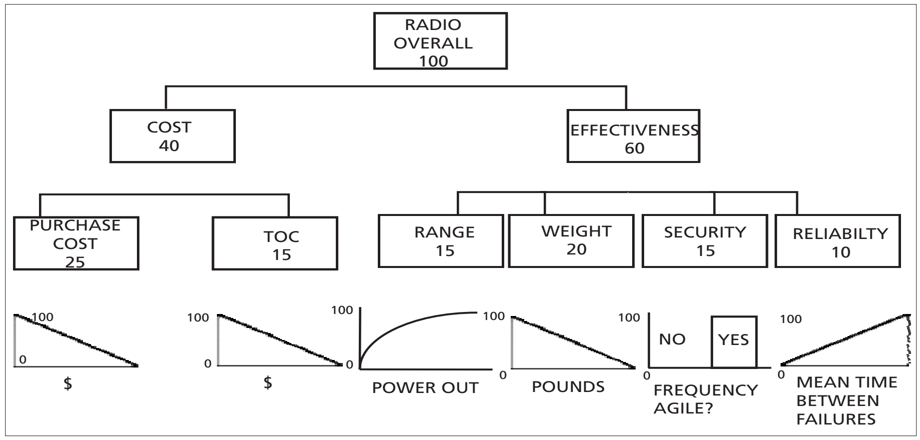


Figure 7-3. Weighted Model for a Portable Radio

		INITIAL COMPARISON			ADJUSTED COMPARISON	
RADIO	**COST (40)**	**EFFECTIVENESS (60)**	**TOTAL (100)**	**COST (50)**	**EFFECTIVENESS (50)**	**TOTAL (100)**
POPIEL 1995	**38**	**20**	**58**	**47.5**	**16.7**	**64.2**
WHAMMO 3000	**18**	**49**	**67**	**22.5**	**40.8**	**63.3**
ZONKER 101	**18**	**46**	**64**	**22.5**	**38.3**	**60.8**

Table 7-2. Radio Model Sensitivity to Cost and Effectiveness Weights.

Thus, we see the importance of choosing our weights carefully and rationally. When we see how fairly small changes in weighting can lead to large changes in outcomes like the shift in model-preferred alternatives between a high-cost, high capability radio to an inexpensive, less capable radio, we must also ask about the validity of the weights in the model. Which is really more important to us—cost or effectiveness? We can also make a strong argument that by building the model before we know the alternatives we are more likely to reflect our organization's values impartially. Further, we can understand why, if we use someone else's model, we need to understand how it works before we accept its results.

We can also use sensitivity analysis to see how much change is necessary in one variable of one alternative to make it the preferred choice—or determine that no amount of change in that area will make it so. Returning to the hypothetical radio scenario (with the original weights in figure 7-3), consider the Zonker Company's situation: it is very competitive with the Whammo model. What can it do to overtake Whammo within the model? The Zonker 101 has achieved maximum performance in three of the four effectiveness criteria, but it scores poorly under Weight. If Zonker can improve performance in this area by lightening a calculable (if they know the shape of the utility curve for weight) number of pounds from the radio, they can achieve a higher score than Whammo. Likewise, they may be able to reduce their profit margin in order to decrease their selling price and become more competitive.

CASE STUDY: THE ANALYSIS PHASE—MODELING USMC MEDIUM-LIFT REQUIREMENTS: THE V-22 OSPREY AND HELICOPTERS

Returning to the V-22 and helicopter analysis, we now examine how the Institute for Defense Analyses evaluated their analytic objective: to compare the V-22 and helicopter alternatives on the basis of cost and operational effectiveness. IDA used a cost-risk-effectiveness approach for their modeling. IDA studied cost separately, and then combined cost and effectiveness to achieve the analytic objective. What little risk they examined was built into their effectiveness measurements. IDA's analytic method is summarized in the diagram below:[6]

Recall that IDA fixed cost by creating two sets of equally expensive aircraft fleets to compare the V-22's and helicopters' effectiveness. The first set was sized to the Marines' desired fleet of 502 V-22s and the second to what DoD was willing to budget for medium lift—356 V-22 cost equivalents.

IDA evaluated the effectiveness of the aircraft fleets in the missions mandated by Congress and in an additional area, Anti-Submarine Warfare, specified to them by DoD. The missions were:

- Amphibious Assault
- Sustained Operations for Logistics Support
- Hostage Rescue/Raid
- Overseas Aircraft Deployment
- Combat Search and Rescue
- Special Operations
- Drug Interdiction
- Anti-Submarine Warfare

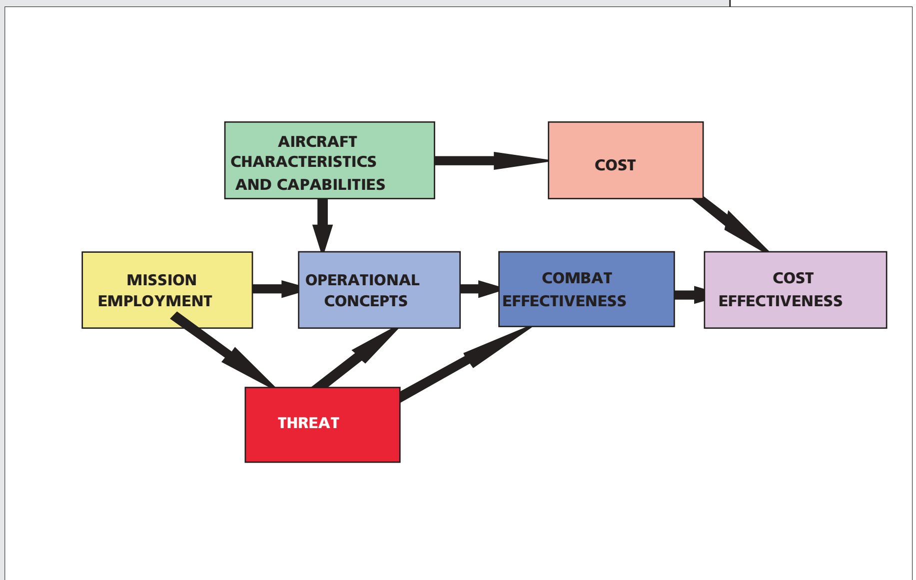
Because the majority of medium-lift aircraft are intended for the amphibious assault role, IDA accorded it particular attention. IDA evaluated the air defense threat in each mission scenario and developed operational concepts that they coordinated with the services and the Joint Staff to ensure they were modeling aircraft employment realistically. Using their abstract operational concepts, IDA estimated the performance of each type of aircraft—the V-22 and the six helicopter options—to determine combat effectiveness. They ran a very large set of excursions to study the fleets' performances in the scenarios.

The Institute for Defense Analyses used at least one model in each of the eight missions. We will concentrate on the amphibious assault scenario because it is the most important to the overall analysis and because IDA used the most complex models for that mission.

IDA used an existing deterministic model to analyze amphibious assaults. This engine was the Amphibious Warfare Model, a 1970's era computer simulation of a conventional theater assault, developed and updated continually by the Center for Naval Analyses. To examine the performance of the options under varying conditions, IDA selected two Department of the Navy case studies and built two corresponding vertical assault forces, each attacking under different battlefield conditions. The assault forces began on amphibious ships 50 nautical miles from the landing zones in both scenarios. A notional Third-World Soviet-style Motorized Rifle Division opposed the Marines in each.

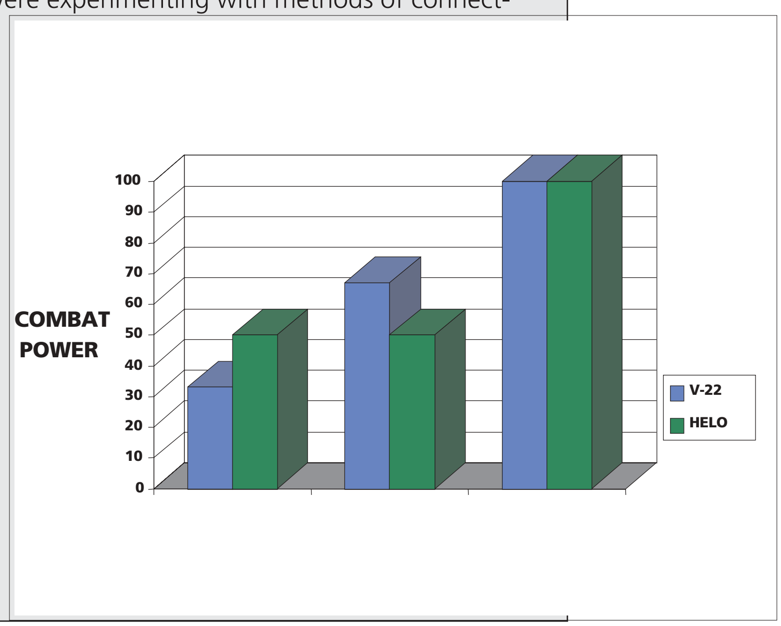
6. Simmons, L.D. Et al, Assessments of Alternatives for the V-22 Assault Aircraft Program, Executive Overview, Institute for Defense Analysis, 1991, pp. 3–4.

IDA made some critical assumptions before running their model. One of the most controversial was a change to Marine Corps doctrine. In 1990, Marine planners assumed the vertical assault would land in only two waves to ensure the first wave had enough combat power to survive enemy reactions until the remainder of the force landed. With the V-22 lower option of 356 aircraft, the Marines could not land 50% of the vertical assault force in the initial wave. IDA assumed that the Marines would accept delivery of the vertical assault combat power in three vice two lifts if the build-up time was not compromised. They reasoned if the Marines' desire was to get a given capability ashore within a time span from H-Hour to time T, the V-22 (with its superior speed) could deliver the same force in the same time frame in three lifts, vice two for the helicopters, and still meet the Marines' requirement, as shown below.



Some critics challenged this assumption, notably Dr. David Chu, Assistant Secretary of Defense for Programs Analysis and Evaluation (see Appendix 3). He noted in his congressional testimony that the 356 aircraft V-22 fleet would have to generate a historically high sortie rate from the assault ships to achieve the build-up in combat power in the time the scenario required. The Marines, however, concurred with IDA's interpretation and skep-tics accused them of redefining doctrine to suit a procurement goal. DoD ultimately decided it was up to the Marines to define their doctrine and if they chose to modify it, that was an internal Marine Corps decision.

A second major assumption the Institute for Defense Analyses made concerned the method the CH-53E heavy lift helicopters, present in each helicopter fleet option, used for lifting external loads slung underneath the aircraft. The Marines were experimenting with methods of connecting two vehicles together as a single, stable load beneath the helicopter to reduce the number of sorties needed during an assault. If they were successful, they would reduce the number of V-22 (and medium helicopter) sorties dedicated to lifting vehicles. At the time of the IDA study, the Marines had not tested these methods at sea. Skeptics were concerned that linking the vehicles would be impractical on darkened, rolling ships and that unlinking them in a landing zone under fire would be too hazardous.



IDA, as in the 1990 Navy study, assumed CH-53Es delivered half the vehicles in dual lifts for smaller assault forces. They assumed all the vehicles would be in dual slings to lift the larger assault forces. IDA did sensitivity analysis to see how the number of vehicles in double slings affected the results of their model; they found the more vehicles that were double-lifted by the CH-53Es, the smaller were the delivery performance differences between medium-lift options, but that the rankings remained the same.

Returning to the assault scenarios, the first study situation was Department of the Navy (DoN) Lift I[7] from 1983. It had the Marines make a night assault in rolling terrain. The low aircraft flight profiles reduced the effectiveness of the enemy air defenses because they were masked by terrain. It also assumed poor reaction times by the defenders. The second scenario was from a 1990 study, DoN Lift II[8]; it set the assault force against a faster-reacting defender about two thirds as well armed as its 1983 counterpart. This time the assault happened in daylight over flat terrain with better fields of fire for the air defenders. Not surprisingly, aircraft casualties were higher in the second case and the casualty differences between aircraft options were larger. IDA ran 388 excursions, varying assault force compositions, tactical factors, threat, and terrain for each aircraft fleet. IDA measured the percentage of the Marine vertical assault element lost attaining a 3:1 advantage in combat power over the defenders to compare and rank the medium-lift options.[9]

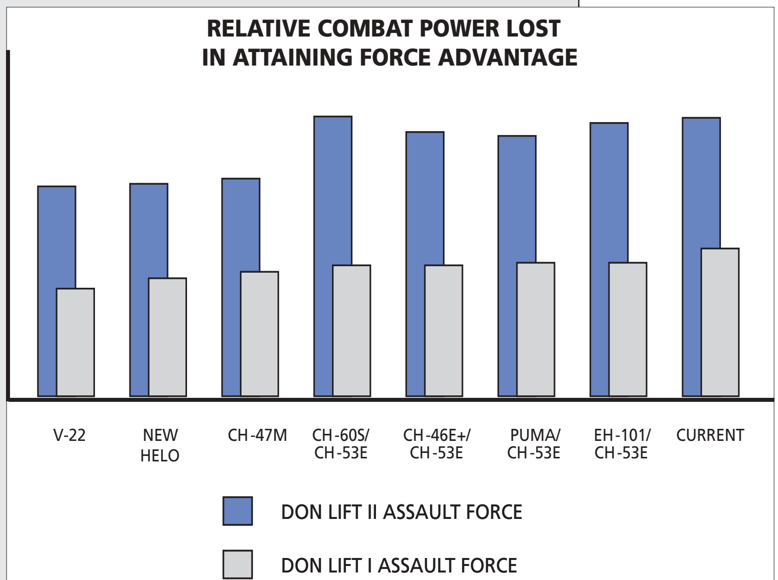
Using both assumptions, IDA ran the model for the aircraft fleets in the two scenarios. With survivability as the principal measure of effectiveness, the V-22 outperformed the helicopters in the amphibious assault mission. They displayed the results in a series of bar graphs, one set for each fleet in each assault case, as shown below (figure 3 from IDA's Executive Summary).

These bar graphs represent the results of the 388 combinations of enemy force composition, tactical factors, threat, and terrain that IDA explored. Those results all fell between the ranges of these bar graphs. In the Amphibious Warfare Model, the size, speed, design, and length of time an aircraft was exposed to enemy air defenses during each possible engagement determined its casualty rate. The V-22, with its higher speed, moved through air defense engagement envelopes faster than the helicopters, therefore it took fewer casualties (although the smaller, harder-to-hit helicopters approached the V-22's survivability rate). Moreover, if DoD opted for the smaller helicopter fleets, they would also have to buy 200 to 260 large, more vulnerable CH-53E helicopters to compensate for the limited external load capability of the smaller helicopters.

Next we will evaluate the validity of the Amphibious Warfare Model for assessing helicopters and the V-22—is this the right model for comparing the medium-lift aircraft alternatives? The level of abstraction of the model for this application is very high because IDA used a very small portion of a very large model for this study. This portion distilled the effectiveness of the aircraft options into a single MOE, survivability, and used a very simple combat engine to evaluate each aircraft. This forces us to ask whether size, speed, length of time in the air defense envelope and the resilience of each aircraft to withstand battle damage are the only important determinants of aircraft effectiveness. How will the V-22 interact with other Marine aircraft for flight operations (flight deck crew turn around time) and for long-range assault (since it can outrun its attack helicopter escort)? Is the number of deck spots important to generate sortie rates? Is unit integrity of the passengers or unloading time important in the landing zone?

7. Department of the Navy Long Term Amphibious Lift Requirement and Optimum Ship Mix Study, Office of the Chief of Naval Operations/Headquarters Marine Corps, 25 May 1983, CONFIDENTIAL.
8. Department of the Navy Integrated Amphibious Operations and USMC Air Support Requirements Study, Office of the Chief of Naval Operations/Headquarters Marine Corps, 5 April 1990, SECRET.
9. IDA ran additional iterations to examine 2.5:1 and 3.5:1 build-ups; the preference rankings of the alternatives remained the same, i.e., the model was not sensitive to how much combat superiority the Marines required.

The behavior of the air defense forces is also greatly simplified; it is based on the air defenders' reaction time. The deployment of the air defenders was arbitrary, but it was the same for all the assault aircraft options in the model. These are all clearly shortcomings in replicating reality; but are they fatal? This is where we must insert professional military judgment to evaluate validity and determine if the relative performance differences of the model carry into the real world. In our opinion (and DoD's, including the Marine Corps), the level of abstraction is appropriate for this 1990 decision and the simplifications were acceptable.



The model is predictive and it needs to be; it is forecasting aircraft casualties during the build up of the assault force in the two cases described above. How accurately we think it predicts depends on our confidence in the assumptions we discussed earlier and our acceptance of this high level of abstraction. We think it will predict the relative behavior of the aircraft alternatives accurately. IDA could make this model more complex—it could incorporate flight operations variables, a more complicated combat engine and more types of air defense weapons. But would these improvements change the outcome of the model output? Probably not.

Now we turn to the reliability of this model—does it model accurately and consistently? We have an inherent reliability problem whenever we rely on contractor projections about aircraft that have not been built yet, in this case the V-22 and the new helicopter. The values for the variables in this model were readily available to IDA from existing databases or were provided directly by the contractors and we have a high level of confidence they reflect real world performance. IDA ran the model hundreds of times and the outcomes were consistent throughout the study. The overall reliability of this model was very high.

This was an important and urgent study; practicality was central to many of IDA's decisions about the model. They knew the six previous studies comparing the V-22 and helicopters had not provided enough information to finally decide this aircraft selection; they felt compelled to add new knowledge to support the decision makers. IDA needed to conserve resources, especially time, producing this analysis. They cleverly adapted existing studies and an existing model to compare the aircraft options, tools previously accepted by the major decision participants. Discussion and controversy quickly focused on the limited number of assumptions and the results of using the model, which was what the participants desired, i.e., they were not distracted examining and debating the model. Enhancing the model to reflect reality in more detail, as described above, was not worthwhile because even if IDA increased the level of detail it would not change the rankings of the aircraft options. The Institute of Defense Analyses scored well in practicality with this study.

IDA presented their findings in six volumes, including the Executive Overview. They presented most of the results in tables and graphs and displayed the utility of the different options, arranged by fleet cost and alternatives. IDA briefed the services, Joint Staff, and Defense Secretariat of their results and eventually testified before Congress.

> Throughout the Analysis Phase, IDA verified they were executing the decision maker's desires. An Office of the Secretary of Defense Steering Committee held five meetings during the course of the study to validate IDA's plan and monitor its progress. Importantly, IDA validated their scenarios with DoD's subject-area experts to include military judgment.

Summary

As our decision becomes more complex, analytic models become less capable of providing clear-cut, definitive answers about how we should choose among alternatives. But even then, well-constructed analytic models provide important insights about how and why the alternatives perform as they do. Combined with professional judgment, this kind of information can guide our choice of courses of action.

Models are important tools that facilitate our decision making by simplifying complex problems, making them easier to understand, change, and manipulate. By using models, we reduce the cost and effort of evaluating alternatives by substituting modified or imaginary environments for actual conditions. Based on the nature of our decision, we select the type analysis we are going to use: exploratory analysis and concept studies for new ideas, cost-risk-effectiveness models for analysis of alternatives, and causal analysis for policy options.

We select or build our models on the basis of the decision we are making, the type of analysis we are doing, and our required levels of abstraction, prediction, and complexity. We prefer to use existing models rather than creating new ones, but we will not force a fit. As with the criteria, we evaluate our models on the basis of validity, reliability, and practicality. We perform sensitivity analysis to identify which variables have the greatest effect on the results of comparing alternatives, enabling us to target changes to options (or the model) to have the greatest effect.

For all their strengths, good models do not guarantee we will make good decisions. Models can have significant shortcomings, especially if they are used incorrectly. Choosing or building the right model to use in a particular decision situation is highly dependent upon the judgment, experience, and collaboration of the decision maker, action officers, and analysts.